Statistical Research Report No. 1.
Institute of Mathematics October 1970
University of Oslo

COMPARISON OF TRANSLATION EXPERIMENTS

by

Erik Nikolai Torgersen
University of Oslo

ABSTRACT

COMPARISON OF TRANSLATION EXPERIMENTS *

BY

Erik Nikolai Torgersen
University of California, Berkeley
University of Oslo

In this paper we treat the problem of comparison of translation experiments. The "convolution divisibility" criterion for "being more informative" by Boll (1955) [2] is generalized to a "ε-convolution divisibility" criterion for ε-deficiency. We also generalize the "convolution divisibility" criterion of V. Strassen (1965) [12] to a criterion for "ε-convolution divisibility". It is shown, provided least favourable "ε-factors" can be found, how the deficiencies actually may be calculated. As an application we determine the increase of information - as measured by the deficiency - contained in an additional number of observations for a few experiments (rectangular, exponential, multivariate normal, one way lay out). Finally we consider the problem of convergence for the pseudo distance introduced by LeCam (1964) [7]. It is shown that convergence for this distance is topologically equivalent to strong convergence of the individual probability measures up to a shift.

---

*) This paper was prepared with the partial support of U.S.Army Grant D-ARO-D31-124-G1135.

COMPARISON OF TRANSLATION EXPERIMENTS

1. Definitions, notations and basic facts.

In [7] LeCam introduced the notion of ε-deficiency of one experiment relative to another. This generalized the concept of "being more informative" which was introduced by Bohnenblust, Shapley and Sherman and may be found in Blackwell [1].

An experiment will here be defined as a pair $\mathcal{E} = ((\chi, \mathcal{A}), (P_\theta : \theta \in \Theta))$ where $(\chi, \mathcal{A})$ is a measurable space and $(P_\theta : \theta \in \Theta)$ is a family of probability measures on $(\chi, \mathcal{A})$. $(\chi, \mathcal{A})$ is the sample space of $\mathcal{E}$ and Θ is the parameter set of $\mathcal{E}$.

Definition.

Let $\mathcal{E} = ((\chi, \mathcal{A}), (P_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (Q_\theta : \theta \in \Theta))$ be two experiments with the same parameter set Θ and let $\theta \rightsquigarrow \varepsilon_\theta$ be a non negative function on Θ.

Then we shall say that $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ if to each decision space $(D, \mathcal{S})$ (i.e. a measurable space) where $\mathcal{S}$ is finite, every bounded loss function $(\theta, d) \rightsquigarrow W_\theta(d)$ on $\Theta \times D$ (W_θ is assumed measurable for each θ) and every risk function r obtainable in $\mathcal{F}$ there is a risk function r' obtainable in $\mathcal{E}$ so that

$$r'(\theta) \leq r(\theta) + \varepsilon_\theta \| W \|, \quad \theta \in \Theta$$

where $\|W\| = \sup_{\theta, d} W_\theta(d)$.

If $\mathcal{E}$ is 0-deficient relative to $\mathcal{F}$ then we shall say that $\mathcal{E}$ is more informative than $\mathcal{F}$ and write this $\mathcal{E} \geq \mathcal{F}$.

If $\mathcal{E} \geq \mathcal{F}$ and $\mathcal{F} \geq \mathcal{E}$, then we shall say that $\mathcal{E}$ and $\mathcal{F}$ are equivalent and write this $\mathcal{E} \sim \mathcal{F}$.

The greatest lower bound of all constants ε such that $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ will be denoted by $\delta(\mathcal{E},\mathcal{F})$ and $\max[\delta(\mathcal{E},\mathcal{F})\delta(\mathcal{F},\mathcal{E})]$ will be denoted by $\Delta(\mathcal{E},\mathcal{F})$.

If $\mathcal{E}, \mathcal{F}$ and $\mathcal{G}$ are experiments then: $0 \leq \delta(\mathcal{E},\mathcal{F}), \delta(\mathcal{E},\mathcal{E}) = 0$ and $\delta(\mathcal{E},\mathcal{F}) \leq \delta(\mathcal{E},\mathcal{G}) + \delta(\mathcal{G},\mathcal{F})$ so that Δ is pseudometric.

Let $\mathcal{E} = ((X,\mathcal{A}),(P_\theta:\theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}),(Q_\theta:\theta \in \Theta))$ be two experiments such that $(P_\theta: \theta \in \Theta)$ is dominated, $\mathcal{Y}$ is a Borel sub set of a complete separable metric space and $\mathcal{B}$ is the class of Borel sub sets of $\mathcal{Y}$. It then follows from theorem 3 in Le Cam's paper [7] that $\mathcal{E}$ is ε-deficient w.r.t $\mathcal{F}$ if and only if there is a Markov kernel M from $(X,\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$ so that :

$$\|P_\theta M - Q_\theta\| \leq \varepsilon_\theta; \quad \theta \in \Theta$$

or equivalently, that there is a sub Markov kernel so that (see section 2)

$$2\|(Q_\theta - P_\theta M)^+\| \leq \varepsilon_\theta; \quad \theta \in \Theta$$

A translation experiment will here be defined as an experiment $\mathcal{E}_P = ((X,\mathcal{A})(P_\theta: \theta \in \Theta))$ where X is a second countable locally compact topological group with Borel class $\mathcal{A}$, $\Theta = X$, P is a probability measure on $\mathcal{A}$ and

$$P_\theta(A) = P(A\theta^{-1}); \; A \in \mathcal{A}, \; \theta \in \Theta$$

Clearly $\mathcal{E}_P$ is uniquely defined by P. μ will always denote a right Haar measure on $(X,\mathcal{A})$.

It will frequently be assumed that P is absolutely continuous i.e. that $P \ll \mu$. This assumption is equivalent with each of the following conditions:

(D_1) $\mathcal{E}_P$ is dominated

(D_2) $(P_\theta : \theta \in \Theta) \sim \mu$

(D_3) $\theta \rightsquigarrow P_\theta(A)$ is continuous for each $A \in \mathcal{A}$

(D_4) $\theta \rightsquigarrow P_\theta$ is strongly continuous.

A dominated translation experiment on the real line is always (this does not hold generally) minimal sufficient, it is not - however - necessarily boundedly complete.

Some of the notations which will be used are:

$(X, \mathcal{A})$ = a measurable space where X is a second countable locally compact group and $\mathcal{A}$ is the class of Borel sub sets of X.

$$\|f\| = \sup_x |f(x)| \quad \text{and} \quad \|\mu\| = \sup\{|\int f(x)\mu(dx)| : \|f\| \leq 1.$$

The convolution $P*Q$ of two probability measures on $\mathcal{A}$ is the measure induced from $P \times Q$ by the map $(x_1, x_2) \rightsquigarrow x_1 x_2$ i.e. $P*Q(A) = P \times Q(\{(x_1,x_2):(x_1 x_2 \in A)\})$.

$C(X)$ = the Banach space of continuous bounded functions on X with sup norm. $M(X)$ is the space of bounded measurable functions.

$C_o(X)$ $(C_o(X)^+)$ is the space of (non negative) continuous functions with compact support.

A subscript θ - with or without affixes - on a probability measure P is the right θ translate of P.

δ_x = the one point distribution in x.

Convergence of probability measures on $C(X)$ is - unless otherwise stated - weak* convergence.

Translation experiments as defined above are - strictly speaking - right translation experiments. Statements on right translation experiments may - by the map $x \rightsquigarrow x^{-1}$ - be translated into statesments for left translation experiments.

2. Some facts on invariance.

It was shown by Boll in [2] that comparison within "invariant pairs" of experiments - under regularily conditions may be based on invariant kernels. In his paper [7] LeCam applied fixed point theorems to ε-comparison within "invariant pairs" of experiments.

We will below summarize a few facts on invariance in the space $\mathcal{X}$ of bounded linear operators from a space $L_1(X,\mathcal{A},\pi)$ to a space $C(\mathcal{Y})^*$. Here $(X,\mathcal{A},\pi)$ is a probability space and $\mathcal{Y}$ is a second countable locally compact Hausdorff space with Borel class $\mathcal{B}$. It may be shown that any T in $\mathcal{X}$ may be represented by a kernel $M: X\times\mathcal{B}\to R$ such that $M(\cdot|x)$ is a measure for each x in X and $M(B|\cdot)$ is measurable for each B in $\mathcal{B}$. M is called a representation of T if $(\mu T)(f) = \int\mu(dx)\int M(dy|x)f(y);\mu\in L_1(\pi),f\in C(\mathcal{Y})$.

Two convex subsets $H \supseteq K$ of $\mathcal{X}$ are of a particular interest in this connection namely :

$$H \overset{\text{definition}}{=\!=} \{T : T \geq 0 \text{ and } \|T\| \leq 1\}$$

$$K \overset{\text{definition}}{=\!=} \{T : T \geq 0 \text{ and } \|\mu T\| = \|\mu\| \text{ when } \mu \geq 0 .\}$$

A kernel M represents an element of H if and only if $M(\cdot|x) \geq 0$ and $\|M(\cdot|x)\| \leq 1$ for almost all x in X. A kernel M which represents an element of H represents an element of K if and only if $\|M(\cdot|x)\| = 1$ for almost all x in X.

Each pair (μ,f) where $\mu\in L_1(\pi)$ and $f\in C_0(\mathcal{Y})$ determines a linear functional : $T \leadsto (\mu T)(f)$ on $\mathcal{X}$.

We choose as a topology for X the smallest topology which makes these functionals continuous. With this topology X becomes a locally convex linear topological space with H as a compact subset. K is compact if and only if $\mathcal{Y}$ is compact.

Let us next consider a measurable group G which acts on X and $\mathcal{Y}$ so that $(x,g) \mapsto g(x)$ and $(y,g) \mapsto g^{-1}(y)$ are both jointly measurable and such that the maps $x \mapsto g(x)$ are Borel equivalences of X and the maps $y \mapsto g(y)$ are homeomorphisms of . We will also assume that $\mu g^{-1} \in L_1(\pi)$ whenever $\mu \in L_1(\pi)$ and $g \in G$.

To each T in X and each g in G we define T^g in X by :

$$(\mu T^g)(f) = ((\mu g^{-1})T)(f \circ g^{-1}); \quad \mu \in L_1(\pi), f \in C_0(\mathcal{Y}).$$

For each g the map $T \mapsto T^g$ is a continuous linear map from X to X which leaves H and K invariant. For each T the map $g \mapsto T^g$ is a homomorphism which is measurable in the sense that $g \mapsto (\mu T^g)(f)$ is measurable whenever $\mu \in L_1(\pi)$ and $f \in C(\mathcal{Y})$. If the kernel M represents T then the kernel M^g: $(x, B) \mapsto M(g(B)|g(x))$ represents T^g. The operator T will be called invariant if $T^g = T$ for each g in G. This is equivalent with the condition that the representing kernel is almost invariant i.e.: $M^g(\cdot|\cdot) = M(\cdot|\cdot)$ a.e. π; $g \in G$ where the exceptional null set is allowed to depend on g.

Before proceeding let us briefly consider the problem of replacing almost invariant kernels with invariant ones. A kernel M will be called invariant if there is an invariant N in with $\pi(N) = 0$ and such that $M^g(\cdot|x) = M(\cdot|x)$ when $x \notin N$. N may be replaced by $\emptyset$ provided there is at least one invariant kernel with the empty set as the exceptional set.

The following considerations on invariant kernels borrows much from C. Boll's paper - they are, however, not entirely the same. We will assume that there is a σ-finite measure - which we always may and will take as a probability measure - τ on G with the property that $\tau(Bg) = 0 \Leftrightarrow \tau(B) = 0$.

Suppose M is almost invariant, let $\mathcal{H}$ be a countable dense sub set of $C_0(\mathcal{Y})$ and let V denote the measurable sub set

$$\{(x,g): \sum_{h\in\mathcal{H}} |M^g(h|x) - M(h|x)| > 0\} \text{ of } \chi \times G.$$

By assumption $\pi(V_g) = 0$ for each g so that - by Fubini's theorem - $\pi(N) = 0$ where $N = \{x: \tau(V_x) > 0\}$. Denote by A the measurable sub set $\{x: M^g(h|x) = \text{constant a.e. } \tau \text{ for each } h \text{ in } \mathcal{H}\} = \{x: M^g(\cdot|x) \text{ is constant in } C(\mathcal{Y})^* \text{ a.e. } \tau\}$.

For any x in A and any g_0 in G we have: (since $\tau(Bg) = 0 \Leftrightarrow \tau(B) = 0$), $M^{gg_0}(\cdot|x)$ is constant a.e. τ $\Rightarrow (M^g)^{g_0}(\cdot|x)$ is constant a.e. $\tau \Rightarrow M^g(\cdot|g_0(x))g_0$ is constant a.e. $\tau \Rightarrow M^g(\cdot|g_0(x))$ is constant a.e. $\tau \Rightarrow g_0(x) \in A$. It follows that $g_0(A) \subseteq A$ and - since g_0 was arbitrary - that $g(A) = A$; $g \in G$. Note that the constants involved in the implications above all are equal to the element $f \hookrightarrow \int M^g(f|x)\tau(dg)$ of $C(\mathcal{Y})^*$.

Suppose next that $x \notin A$. Then $\sum_{h\in\mathcal{H}} |M^g(h|x)-M(h|x)| > 0$ on a set of positive τ measure and this implies that $x \in N$.

Hence $\pi(A) = 1$. We may now modify M to an invariant $\hat{M}$ by writing

$$\hat{M}(f|x) = \int M^g(f|x)\tau(dg);\ f \in C(\mathcal{Y});\quad x \in A$$

$\hat{M}$ is equivalent with M since $x \notin N$ implies $x \in A$ and $\hat{M}(f|x) = M(f|x)$. $\hat{M}$ is invariant on A since - for any $x \in A$ and $f \in C_0(\mathcal{Y})$ - $\hat{M}^{g_0}(f|x) = \hat{M}(f \circ g_0^{-1}|g_0(x)) = \int M^g(f \circ g_0^{-1}|g_0(x))\tau(dg)$ $= \int M^{gg_0}(f|x)\tau(dg) = \int M^g(f|x)\tau(dg) = \hat{M}(f|x)$.

We return to the considerations of invariant operators. The situation we will encounter in this paper involves a set Θ, for each g in G a function $\theta \rightsquigarrow g(\theta)$ from Θ to Θ and finally an invariant family of triples $\{(P_\theta, Q_\theta, \varepsilon_\theta);\ \theta \in \Theta\}$ where the P_θ's are probability measures in $L_1(\pi)$, the Q_θ's are probability measures in $C(\mathcal{Y})^*$ and the ε_θ's are non negative real numbers. That the family is invariant means that $P_\theta g^{-1} = P_{g(\theta)}$; $Q_\theta g^{-1} = Q_{g(\theta)}$ and $\varepsilon_\theta = \varepsilon_{g(\theta)}$ for any θ in Θ and any g in G.

Two convex and invariant sets of interest here are

$$H_0 = \{T : T \in H;\ \|Q_\theta - P_\theta T\| + 1 - (P_\theta T)(\mathcal{Y}) = 2\|(Q_\theta - P_\theta T)^+\| \leq \varepsilon_\theta!\ \theta \in \Theta\}.$$

and

$$K_0 = H_0 \wedge K.$$

H_0 is compact, but may - of course - be empty. If, however, $T \in H_0$ is represented by M, then any kernel of the form $(x,B) \rightsquigarrow M(B|x) + (1-M(\mathcal{Y}|x))S(B|x)$ where S represents an element of K defines an element of K_0. Moreover if T in H_0 is invariant then this operator is invariant provided S is almost invariant. We may therefore restrict our attention to H_o.

It follows directly from theorem 1 in Day's paper [4] and part (a) of section 4 in the same paper that there is an invariant T in H_o provided $H_o \neq \emptyset$ and provided there is at least one invariant mean (left or right) on the class of bounded measurable functions on G. A general reference on invariant means on topological groups is Greenleaf [5]. It follows from the Kakutani - Markov fixed point theorem that there are invariant means on any abelian group.

The condition of the existence of invariant means is equivalent with the possibly more familiar requirement that there is a net $\{\lambda_t\}$ of probability measures on G (which always may be chosen so that they have finite supports) which is asymptotically right invariant in the sense that $\lambda_t(Bg) - \lambda_t(B) \to 0$ for each measurable sub set B of G and each g in G. A proof of the existence of fixed points in H_o when $H_o \neq \emptyset$ may be based upon this fact, using the same type of arguments as in the proof of Hunt, Stein's theorem in [9]. (This was the approach used by Boll in [2]). Suppose finnally that G is a second countable locally compact topological group with the Borel class as the class of measurable sets. Then it may be shown (Theorem 3.6.2 in [5]) that there is an invariant mean if and only if there is an expanding sequence $\{C_N\}$ of compact sets converging to G and having the property that the corresponding normalized restrictions $\{\lambda_n\}$ of a left Haar measure are strongly convergent to left invariance i.e.: $\sup_B[\lambda_n(gB)-\lambda_n(B)] \to 0$ as $n \to \infty$ for each g in G.

<u>Example .</u> Let $\mathcal{E}$ be the experiment

$((X^n, \mathcal{A}^n);\ (P_\theta : \theta \in \Theta))$ where $X = \{ \cdots -1, 0, 1, \cdots \}$,

$\mathcal{A}$ is the class of all sub sets of X and

$P_\theta(A) = P(A-(\theta,\theta,\cdots,\theta));\ A \in \mathcal{A},\ \ \theta \in \Theta$. We want to compare this experiment with the "total information" experiment $\mathcal{F}$ which may be described as above with $n = 1$ and $P = \delta$, where δ is the one point distribution in 0. The deficiency $\delta(\mathcal{E},\mathcal{F})$ is then equal to $\inf_M \sup_\theta \|P_\theta M - \delta_\theta\|$. By invariance it suffices to consider invariant kernels M from X^n to X so that $\delta(\mathcal{E},\mathcal{F}) =$

$= \inf_M \|PM-\delta\| = 2(1-\sup_M (P^n M)(0)) = 2(1-\sup_\gamma \sum \gamma(x_1,\cdots,x_n)P(x_1,\cdots,x_n))$,

where γ is and may be any function from X^n to $[0,1]$ such that

$$\sum_x \gamma(x_1+x, x_2+x, \cdots, x_n+x) = 1;\quad x_1, x_2, \cdots x_n \in X.$$

It follows that

$$\delta(\mathcal{E},\mathcal{F}) = 2(1- \sum_{x_2,\cdots,x_n} \sup_x P(x, x+x_2, \cdots, x+x_n))$$

As a particular case, consider the situation where P is the distribution of a sample of size n from the uniform distribution on $V = \{a+d,\ a+2d, \cdots, a+Nd\}$ where $a, d \neq 0$ and $N > 0$ are integers. Then $\bigvee_x P(x, x+x_2, \cdots, x+x_n)$ is positive if and only if $(x_2,\cdots,x_n) \in V^{n-1}$-diagonal$(V^{n-1})$ and then

$\bigvee_x P(x, x+x_2, \cdots, x+x_n) = N^{-n}$. It remains therefore to find the total number of distinct $(n-1)$ tuples $(x_2,\cdots,x_n)$ which are of the form $d(k_2,\cdots,k_n) - d(k\ ,\cdots,k)$ where $k, k_2, \cdots, k_n \in \{1,2,\cdots,N\}$. Putting $A_k = V^{n-1} - d(k\ ,\cdots,k);\ k = 1,2,\cdots,N$ this number may be written $\# (A_1 \cup \cdots \cup A_N) = \sum_i \#(A_i) - \sum_{i<j} \#(A_i \cap A_j) + \cdots \pm \#(A_1 \cap \cdots \cap A_N)$.

Now $\#(A_{s_1} \cap \cdots \cap A_{s_k}) = (N-(s_k - s_1))^{n-1}$ when $s_1 < s_2 < \cdots < s_k$.

Hence

$$\#(A_1 \cup \cdots \cup A_N) = NN^{n-1} + \sum_{k=2}^{N} (-1)^{k-1} \sum_{s_k - s_1 \geq k-1} \binom{s_k - s_1 - 1}{k-2} (N-(s_k - s_1))^{n-1}$$

$$= N^n + \sum_{k=2}^{N} (-1)^{k-1} \sum_{i=k-1}^{N-1} \binom{i-1}{k-2} (N-i)^n = N^n - (N-1)^n \ .$$

It follows that

$$\delta(\mathcal{E}, \mathcal{F}) = 2(1 - \tfrac{1}{N})^n \ .$$

3. ε-deficiency and ε-factorization.

It was shown by C. Boll in [2] (for the case $X = R^k$) that a dominated translation experiment $\mathcal{E}_P$ is more informative than $\mathcal{E}_Q$ if and only if P is a factor in Q for convolution, i.e. that there exists a probability measure M so that $Q = M * P$ This implies that most factorization theorems for probability distributions have "experiment interpretations". For example - if Q is normal then a dominated experiment $\mathcal{E}_P$ is more informative than $\mathcal{E}_Q$ if and only if P is a normal distribution with smaller variance.

The result above (as shown by Boll and theorem 1 below) is immediate consequence of the simple fact that a Markov kernel M from X to X is translation invariant if and only if the auxiliary experiment defined by M is a translation experiment.

Theorem 1.

Let P and Q be probability measures on $(X, \mathcal{A})$ and let $\varepsilon \geq 0$ be a constant.

(i) If there is a probability measure M on $\mathcal{A}$ so that

$\|M * P - Q\| \leq \varepsilon$ then $\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \varepsilon$

(ii) Suppose $M(X)$ has an invariant mean and that $P \ll \mu$. Then $\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \varepsilon$ if and only if there is a probability measure M on $\mathcal{A}$ so that $\|M * P - Q\| \leq \varepsilon$.

Proof.

(i) Define the Markov kernel N from X to X by

$N(A|x) = M(A\,x^{-1})$; $A \in \mathcal{A}$, $x \in X$.

Then

$(P_\theta N)(A) = \int N(A|x)P_\theta(dx) = (M * P_\theta)(A)$

Hence

$\| P_\theta N - Q_\theta \| = \| M * P - Q \| \leq \varepsilon$

(ii) Suppose $\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \varepsilon$. By the invariance considerations

i in section 2 there is an invariant Markov kernel N so that

$\| NP - Q \| \leq \varepsilon$

Define M on $\mathcal{A}$ by

$M(A) = N(A|0); \quad A \in \mathcal{A}$

Then - by invariance:

$N(A|x) = M(A\, x^{-1}); \quad A \in \mathcal{A}, \quad x \in X$.

so that

$Np = M * P$

Corollary 2.

$\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \inf_M \| M * P - Q \|$ and "=" holds if $P \ll \mu$ and $M(X)$ has an invariant mean.

Consider now the testing problem: $(P_\theta : \theta \in \Theta)$ against $\{Q\}$. For any prior M and any $\alpha \in [0,1]$ let $\beta_M(\alpha)$ be the power of the most powerful level α test for testing $M * P$ against Q. Suppose M_0 is a least favourable prior distribution on Θ for all significance levels α. Then :

$$\beta_{M_O}(\alpha) \leq \beta_M(\alpha) \quad \text{for all} \quad \alpha \in [0,1]$$

so that

$$\|M_o * P - Q\| = 2\sup_\alpha(\beta_{M_O}(\alpha) - \alpha) \leq 2\sup_\alpha(\beta_M(\alpha) - \alpha) =$$

$$\|M * P - Q\| \quad \text{for all} \quad M.$$

In particular , if $P << \mu$ and $M(X)$ has an invariant mean, then $\delta(\mathcal{E}_P,\mathcal{E}_Q) = \|M_o * P - Q\|$.

Let P, Q and M be probability measures on $\mathcal{A}$ and $\varepsilon \geq 0$ a constant so that

$$\|M * P - Q\| \leq \varepsilon$$

Let f be a bounded measurable function on $(X, \mathcal{A})$. Then :

$$\int f dQ = \int f(x)(M * P)(dx) + \int f(x)(Q - M * P)(dx) \leq$$

$$\int\left[\int f(xy)P(dy)\right]M(dx) + \varepsilon\|f\|$$

so that

$$\int f dQ \leq \sup_x \int f(xy)P(dy) + \varepsilon\|f\|$$

In [12] Strassen proved that this inequality for $\varepsilon = 0$ and all $f \in C(\quad)$ implied the existence of a probability M so that

$$(3.3) \qquad M * P = Q$$

The next theorem generalizes this to an arbitrary $\varepsilon \geq 0$.

Theorem 3.

Let P and Q be probability measures on $(X, \mathcal{A})$ and $\varepsilon \geq 0$ a constant. Then there exists a probability measure M so that

$$\|M * P - Q\| \leq \varepsilon$$

if and only if

$$\int f dQ \leq \sup_x \int f(xy) P(dy) + \varepsilon \|f\| \quad , \qquad f \in C(X).$$

Proof.

It remains to prove the "if".

For each $f \in C(X)$ put

$$\Phi(f) = \sup_x \int f(xy) P(dy)$$

and

$$\psi(f) = \varepsilon \| f \|$$

Then Φ and ψ are strongly continuous sub linear functionals on $C(X)$. By theorem 10 in [12]

$$\{M * P: M \text{ is a probability measure on } \mathcal{A}\} = \{\lambda: \lambda \in C(X)^* \; , \quad \lambda \leq \Phi\}$$

and it may be checked that

$$\{\lambda: \quad \lambda \in C(X)^* \; , \quad \|\lambda\| \leq \varepsilon\} = \{\lambda: \quad \lambda \leq \psi\} \; .$$

By assumption

$$\int f dQ \leq \psi(f) + \Phi(f); \qquad f \in C(X).$$

By theorem 1 in [12] there are linear functionals $\lambda_1 \leq \Phi$ and $\lambda_2 \leq \psi$ in $C(X)^*$ so that

$$\int f dQ = \lambda_1(f) + \lambda_2(f)$$

It follows that there is a probability measure M so that:

$$\|Q - M * P\| \leq \varepsilon$$

$\square$

Remark

Since $\|2f - \|f\| \; \| \leq \|f\|$ when $f \geq 0$ the condition could be replaced by $\int f dQ \leq \sup_x \int f(xy) P(dy) + \frac{\varepsilon}{2} \|f\|$, $f \geq 0$, $f \in C_0(X)^+$.

By the replacement $f \rightsquigarrow -f$ this condition may also be written

$$\inf_x \int f(xy)P(dy) \leqq Q(f) + \varepsilon\|f\| \; ; \qquad f \in C(X)$$

or equivalently

$$\inf_x \int f(xy)P(dy) \leqq \inf_x \int f(xy)Q(dy) + \varepsilon\|f\| ; \qquad f \in C(X)$$

so that

$$\inf_M \|M*P-Q\| = \sup_{\|f\| \leqq 1} \left[\inf_x \int f(xy)P(dy) - \inf_x \int f(xy)Q(dy)\right] ; f \in C(X).$$

These relations may be interpreted as relations between risks for invariant procedures, since - for example - $\int f(xy)P(dy)$ is the (constant) risk for the invariant procedure $y \rightsquigarrow xy$ w.r.t. the decision problem $(X, \mathcal{A}, (\theta,t) \rightsquigarrow f(t\theta^{-1}))$.

Let f be a bounded measurable function on X, let P be any probability measure on $\mathcal{A}$ and let us suppose that $M(X)$ has an invariant mean. Let $\{\lambda_n\}$ be a sequence of probability measures on $\mathcal{A}$ which is strongly convergent to left invariance and let W be the loss function $(\theta,t) \rightsquigarrow f(t\theta^{-1})$ on $(\mathcal{X}, \mathcal{A})$. Then we have :

$$\int (P_\theta \rho W_\theta)\lambda_n(d\theta) = \int P(dx)\int \lambda_n(d\theta)\int \rho(dt/x\theta) f(t\theta^{-1}) =$$

$$= \int P(dx)\int \lambda_n(x^{-1}d\theta)\int \rho(dt/\theta) f(t\theta^{-1}x)$$

$$= \int P(dx)\int \lambda_n(d\theta)\int \rho(dt/\theta) f(t\theta^{-1}x) + \int r_n(x)P(dx)$$

where $r_n(x) = \int \lambda_n(x^{-1}(d\theta))\int \rho(dt/\theta) f(t\theta^{-1}x) - \int \lambda_n(d\theta)\int \rho(dt/\theta) f(t\theta^{-1}x)$

so that $|r_n(x)| \leqq \|\lambda_n(x^{-1}(\cdot))-\lambda_n\| \; \|f\|$.

Hence

$$\int(P_\theta \rho W_\theta)\lambda_n(d\theta) = \int\lambda_n(d\theta)\int P(dx)\int\rho(dt/\theta)f(t\theta^{-1}x)+\int r_n(x)P(dx)$$

$$= \int\lambda_n(d\theta)\int\rho(dt/\theta)\int P(dx)f(\,t\theta^{-1}x)+\int r_n(x)P(dx)$$

$$\geqq \inf_x\int f(xy)P(dy)-\int\|\lambda_n(x^{-1}(\cdot))-\lambda_n\| \; \|f\|P(dx)$$

$$\to \inf_x\int f(xy)P(dy) \quad \text{as} \quad n\to\infty .$$

It follows that $\inf_x\int f(xy)P(dy)$ is the value of the statistical game - and is therefore the minimax value. If we combine this with theorem 3 we get the following generalization of theorem 1 part (ii) :

(ii') Suppose that $M(X)$ has an invariant mean. Then there is a probability measure M so that $\|M*P - Q\| \leqq \varepsilon$ if and only if $\inf_\rho \sup_\theta P_\theta\rho W_\theta \leqq Q_\theta\sigma W_\theta + \varepsilon\|W_\theta\|$ for any loss function $(\theta,t) \hookrightarrow f(t\theta^{-1})$ where $f \in C_0(X)$ (f is bounded measurable) and the inf is taken over all decision procedures ρ. (In particular this holds if there is a Markov kernel M so that $\|P_\theta M - Q_\theta\| \leqq \varepsilon; \quad \theta\in\Theta$).

We introduce now the notations :

$$\delta(P,Q) = \inf_M \| M*P-Q \| = \min_M \| M*P-Q \|$$

$$= \sup_{\|f\| \leq 1} (\inf_x \int f(xy)P(dy) - \inf_x \int f(xy)Q(dy))$$

$$= \sup_{\|f\| \leq 1} (\inf_x \int f(xy)P(dy) - Q(f))$$

$$\Delta(P,Q) = \delta(P,Q) \vee \delta(Q,P) =$$

$$= \sup_{\|f\| \leq 1} |\inf_x \int f(xy)P(dy) - \inf_x \int f(xy)Q(dy)|$$

Then $\delta(P,Q) = \delta(\mathcal{E}_P, \mathcal{E}_Q)$ $(\Delta(P,Q) = \Delta(\mathcal{E}_P, \mathcal{E}_Q))$ provided P is (P and Q are) absolutely continuous. In any case δ and Δ have decision theoretical interpretations implied by the above inequalities. It is clear that :

$$\delta(P,P) = 0, \quad 0 \leq \delta(P,Q) \leq 2, \quad \delta(P,Q) \leq \delta(P,R) + \delta(R,Q)$$

and that Δ is a pseudometric.

It follows from theorem 5.2 in chapter 5 in Parthasarathy's book [10] that - in the abelian case - $\Delta(P,Q) = 0$ if an only if P is a shift of Q. It is trivial that $\Delta(P,Q) = 0$ when P is a left shift of Q. H. Heyer has [6] made me aware of the fact that this condition is necessary in the non-commutative case also. This may be demonstrated as follows: Let P and Q be probability measures on $\mathcal{A}$. If there are probability measures R and S

such that $P = R*Q$ and $Q = S*P$ (i.e. $\Delta(P,Q) = 0$) then we will show that P is a left shift of Q (i.e. R and S may be chosen as one point distributions). We may - without loss of generality - assume that the identity is a support point for both R and S. Put $V = R*S$ and $W = S*R$. Then $P = V*P$ and $Q = W*Q$ so that $P = (\frac{1}{n}\sum_{i=1}^{n} V^i)*P$ and $Q = (\frac{1}{n}\sum_{i=1}^{n} W^i)Q$. It follows that (see theorem 2.1 in chapter 3 in [10]) that the sequences $\frac{1}{n}\sum_{i=1}^{n} V^i$; $n = 1,2,\cdots$ and $\frac{1}{n}\sum_{i=1}^{n} W^i$; $n = 1,2,$ are tight.

Let $\bar{V}$ and $\bar{W}$ be cluster points of respectively the first and the latter sequence. Then $\bar{V}$ and $\bar{W}$ are idempotent (i.e. $\bar{V}*\bar{V} = \bar{V}$ and $\bar{W}*\bar{W} = \bar{W}$), $\bar{V}*V = V*\bar{V} = \bar{V}$, $\bar{W}*W = W*\bar{W} = \bar{W}$, $P = \bar{V}*P$ and $Q = \bar{W}*Q$. By theorem 3.1 in chapter 3 in [10], $\bar{V}$ and $\bar{W}$ are Haar measures on compact subgroups of X. Since the identity is a support point for R as well as for S these relations implies that the supports of R and S are contained in the supports of $\bar{V}$ and $\bar{W}$. Hence $W*\bar{V} = S*R*\bar{V} = \bar{V}$ and $\bar{W}*V = \bar{W}*R*S = \bar{W}$ so that $\bar{V} = \bar{W}*\bar{V} = \bar{W}$. It follows that $P = \bar{V}*P = \bar{V}*S*P = \bar{V}*Q = \bar{W}*Q = Q$.

4. Product experiments, examples.

One would expect that reasonable measures of information could be constructed from the deficiencies. In the case of dichotomies - for example - we could use the deficiency of the minimum informative dichotomy w.r.t. the dichotomies $\mathcal{G}^n$ as a measure of the information contained in $\mathcal{G}^n$. This measure behaves nicely as a function of n and is equivalent with the Bayes risk for the uniform prior [3] ,[14]. Alternatively one could consider the deficiency of $\mathcal{G}^n$ w.r.t. the total informative experiment.

Neither of these constructions yield useful results for translation experiments. To get an idea of the difficulties involved, consider a dominated translation experiment $\mathcal{E}_P$ on the additive group of real numbers. Let $\mathcal{M}_i$ and $\mathcal{M}_a$ denote the experiments $((\{0\},\{\emptyset,\{0\}\}),\ (D;\ \theta \in R))$ and $\mathcal{M}_a = \mathcal{E}_\delta$ respectively (δ is the one-point distribution in 0). Then - by the Markov kernel criterion:

$$\mathcal{M}_i \leq \mathcal{F} \leq \mathcal{M}_a$$

for any experiment $\mathcal{F} = ((\mathcal{Y},\mathcal{B}),\ (Q_\theta:\ \theta \in \Theta))$ such that $\theta \rightsquigarrow Q_\theta(B)$ is measurable for each $B \in \mathcal{B}$.

By the Markov kernel criterion: $\delta(\mathcal{M}_i, \mathcal{E}_P{}^n) = \inf\limits_M \sup\limits_\theta \|M - P_\theta{}^n\|$.

Let S be the map $(x_1,\dots,x_n) \rightsquigarrow \sum\limits_{i=1}^{n} x_i$. Then

$$\|M - P_\theta{}^n\| \geq \|MS^{-1} - P_\theta{}^{n*}\| = \|MS^{-1} - P_{n\theta}{}^{n*}\| \to 2 \quad \text{as} \quad \theta \to \infty$$

By the Markov kernel criterion again: $\delta(\mathcal{E}_P{}^n, \mathcal{M}_a) =$ $\inf\limits_M \sup\limits_\theta \|MP_\theta{}^n - \delta_\theta\|$.

Let the additive group G of real numbers act on χ, χ^n and Θ so that $g(x) = x+g$, $g(x_1,\cdots,x_n) = (x_1+g,\cdots,x_n+g)$. $g(\theta) = \theta+g$; $x \in \chi$, $(x_1,\cdots,x_n) \in \chi^n$ and $\theta \in \Theta$. It follows by invariance that we may restrict attention to invariant kernels.

Let M be an invariant kernel. Then :

$$\sup_\Theta \|MP_\theta^{\,n} - \delta_\theta\| = \|MP^n - \delta\| = 2(1-MP^n(\{0\})) =$$

$$= 2-2\int M(\{0\}|x_1,\cdots,x_n)f(x_1)\cdots f(x_n)dx_1\cdots dx_n$$

$$= 2-\int 2M(\{0\}|y_1,y_1+y_2,\cdots,y_1+y_n)f(y_1)f(y_1+y_2)\cdots f(y_1+y_n)dy_1\cdots dy_n$$

$$= 2-2\int[\int M(\{-y_1\}|0,y_2,\cdots,y_n)f(y_1)f(y_1+y_2),\cdots,f(y_1+y_n)dy_1]dy_2\cdots dy_n$$

$$= 2$$

since there is, for each $(y_2,\cdots,y_n)$, at most a countable number of numbers y_1 so that $M(\{-y_1\}|0,y_2,\cdots,y_n) > 0$.
Hence

$$\delta(\mathscr{E}_P^{\,n}, \mathscr{M}_a) = 2 .$$

It follows that the deficiencies $\delta(\mathscr{M}_i, \mathscr{E}^n)$ and $\delta(\mathscr{E}^n, \mathscr{M}_a)$ are useless as measures of information in the translation case. Deficiencies may, however, be used to study the amount of information contained in an additional number of observations. To see this, let us consider a few examples of semigroups (w.r.t. experiment multiplication) of translation experiments. We will use the notation $a_n \sim b_n$ to indicate that $\lim_{n\to\infty} a_n/b_n = 1$. The χ^2 distribution with n degrees of freedom (and its distribution function) will be denoted by Γ_n.

Example 4.

(Rectangular distribution, unknown scale parameter). Let $X =]0,\infty[$ with multiplication as group operation. Let R be the rectangular distribution on $]0,1]$. Then R_θ is the rectangular distribution on $]0,\theta]$. Consider the experiments $\mathscr{E}_R^n$, $n = 1,2,\cdots$.

By sufficiency $\mathscr{E}_R^n \sim \mathscr{E}_{P_n}$ where

$$dP_n/dR = nx^{n-1}I_{]0,1]}(x).$$

Consider the problem of testing :

$$P_{n,\theta}: \theta > 0 \quad \text{against} \quad P_{n+a} \quad \text{where} \quad a > 0.$$

It is then easily checked that least favourable distribution assigns mass 1 in $\theta = 1$. Hence

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+a}) = \int |nx^{n-1} - (n+a)x^{n+a-1}|\,dx = 2(1+\tfrac{a}{n})^{-\frac{n}{a}}\left[\frac{a}{n+a}\right] .$$

It follows that $\delta(\mathscr{E}^n, \mathscr{E}^{n+a}) \sim \frac{2}{e}\frac{a}{n} = (0.73\ldots)\frac{a}{n}$.

Example 5.

Let $T_1, T_2, \cdots$ be independently and identically distributed, each having the density $\lambda e^{-\lambda t}$; $t > 0$ where $\lambda > 0$ is unknown. Let $\mathscr{F}^n$ be the experiment obtained by observing $T_1, T_2, \cdots, T_n$. Then - by sufficiency - $\mathscr{F}^n$ is equivalent with the translation experiment $\mathscr{E}_{\Gamma_{2n}}$ on $(]0,\infty[\cdot)$. We will see in the next example that

$$\delta(\mathscr{F}^n, \mathscr{F}^{n+a}) =$$

$$= \|\Gamma_{2n+2a} - \Gamma_{2n,1+\frac{a}{n}}\| \sim \sqrt{\frac{2}{\pi e}}\;\frac{a}{n} = (0.48\ldots)\,\frac{a}{n} .$$

Example 6.

Consider the translation experiments $\mathscr{E}_{\Gamma_n}$; $n = 1,2,\ldots$ on $(]0,\infty[,\cdot)$. For testing $\{\Gamma_{n,\theta}\}$ against Γ_{n+a} the one point distribution in $(1+\frac{a}{n})$ is least favourable at all levels. Hence:

$$\delta(\mathscr{E}_{\Gamma_n},\mathscr{E}_{\Gamma_{n+a}}) = \|\Gamma_{n+a} - \Gamma_{n,1+\frac{a}{n}}\| = E\left|1 - (1+\tfrac{a}{n})\frac{\Gamma'_{n+a}((1+\frac{a}{n})X)}{\Gamma'_n(X)}\right|$$

where X is a random variable which is χ^2 distributed with n degrees of freedom. If we introduce the random variable $Z = \frac{X-n}{\sqrt{2n}}$, then - by Stirling's approximations and the asymptotic normality of Z - this implies :

$$\delta(\mathscr{E}_{\Gamma_n},\mathscr{E}_{\Gamma_{n+a}}) \sim \frac{a}{2n} E|U^2 - 1|$$

where U is standard normal. Now :

$$E|U^2 - 1| = 2(\Gamma_1(1) - \Gamma_3(1)) = 2\sqrt{\frac{2}{\pi e}} \quad \text{since} \quad \Gamma_3(1) = \Gamma_1(1)+\sqrt{\frac{2}{\pi e}}\ .$$

It follows that :

$$\delta(\mathscr{E}_{\Gamma_n},\mathscr{E}_{\Gamma_{n+a}}) \sim \sqrt{\frac{2}{\pi e}}\,\frac{a}{n} = (0.48\ldots)\,\frac{a}{n}\ .$$

Example 7.

Let $X = (X_1,\cdots,X_k)'$ be multivariate normal with unknown mean $\xi = (\xi_1,\cdots,\xi_k)'$ and known positive definite covariance matrix D. Let $\mathscr{E}^n$ be the experiment consisting of n independent observations of X. Then $\delta(\mathscr{E}^n,\mathscr{E}^{n+a}) \sim 2k\Gamma'_k(k)\,\frac{a}{n}$ as $n \to \infty$. We may - without loss of generality - assume that D is the identity matrix - and then this is a particular case of the next example.

Example 8.

Let X be as in the previous example, and let $\mathcal{H}^{n_1,n_2,\cdots,n_k}$ denote the experiment obtained by taking n_i observations on X_i; $i = 1,2,\cdots,k$. Then $\mathcal{H}^{n_1,\cdots,n_k}$ is equivalent with $\mathcal{E}_{P_{n_1,n_2,\cdots,n_k}}$ on $(R^k,+)$ where $P_{n_1,\cdots,n_k}$ is the joint distribution of k independent normally distributed variables $Z_1,\cdots,Z_k$ where $EZ_i = 0$ and $\mathrm{Var}Z_i = 1/n_i$. Let $a_i > 0$; $i = 1,2,\cdots,k$. The least favourable distribution for testing $P_{n_1,n_2,\cdots,n_k}$, ; $\xi \in R^k$ against $P_{n_1+a_1,n_2+a_2,\cdots,n_k+a_k}$ is the one point distribution in $(0,0,\cdots,0)$ so that

$$\delta(\mathcal{H}^{n_1,\cdots,n_k}, \mathcal{H}^{n_1+a_1,\cdots,n_k+a_k}) = E\left|1 - \prod_{i=1}^{k} \left(\sqrt{1+\frac{a_i}{n_i}}\, e^{-\frac{1}{2}\frac{a_i Y_i^2}{n_i}}\right)^2\right|$$

where $Y_1,Y_2,\cdots,Y_k$ are independent standard normal variables. Put $N = \sum_{i=1}^{k} n_i$ and suppose $n_i/N \to \lambda_i > 0$ as $N \to \infty$. Then

$$\delta(\mathcal{H}^{n_1,\cdots,n_k}, \mathcal{H}^{n_1+a_1,\cdots,n_k+a_k}) \sim E\left|\sum_{i=1}^{k} a_i/\lambda_i (Y_i^2 - 1)\right|(2N)^{-1}$$

as $N \to \infty$. In particular, if $a_i/\lambda_i = C$; $i = 1,2,\cdots,k$, then

$$\delta_N \sim \frac{C\, E\left|\sum_{i=1}^{k} Y_i^2 - k\right|}{2N} = \frac{2kC(\Gamma_k(k)-\Gamma_{k+2}(k))}{2N} \qquad \text{as } N \to \infty$$

Since $\Gamma_k(k)-\Gamma_{k+2}(k) = 2\Gamma_k'(k)$ this may be written

$\delta_N \sim \dfrac{2Ck\Gamma_k'(k)}{N}$ as $N \to \infty$. The asymptotic equivalence in example 7 follows now by putting $C = ka$ and $N = kn$.

The asymptotic behavior of the powers of the experiments in these examples is very different from the asymptotic behavior of powers of experiments with finite parameter sets. If $\mathscr{E} = ((\mathscr{Y},\mathscr{B}); (Q_\theta : \theta \in \Theta))$ is an experiment with finite parameter set Θ, then $\delta(\mathscr{E}^n, \mathscr{M}_a) \to 0$ as $n \to \infty$ with exponential speed - provided $\theta \rightsquigarrow Q_\theta$ is 1-1, and $\mathscr{E} \dagger \mathscr{M}_a$.

5. Convergence of translation experiments.

When is $\mathcal{E}_{P_1}, \mathcal{E}_{P_2}, \cdots$ asymptotically equivalent with $\mathcal{E}_P$? A sufficient condition is clearly strong convergence of $P_1, P_2, \cdots$ to P up to a shift. The necessity - under regular conditions - of this condition is the essential part of the following theorem.

Theorem 9.

Let P be absolutely continuous. Then $\Delta(P_n, P) \to 0$ if and only if there exist elements $a_1, a_2, \cdots$ in X so that $\|\delta_{a_n} * P_n - P\| \to 0$.

Remark: This result is analogous to the result [14] that a net of experiments on a finite and fixed sample space converges to a minimal sufficient experiment if and only if the indivudual probability measures converge up to permutations of the sample space. This theorem could have been formulated for nets as well, since it merely states that two pseudometrics on the set of absolutely continuous probability measures are topologically equivalent - the other pseudometric being: $(P,Q) \hookrightarrow \inf_a \|\delta_a * P - Q\|$. The proof follows directly from proposition 10 and proposition 11 below.

Proposition 10.

Let X be a second countable locally compact group with Borel class $\mathcal{A}$. Let N_n; $n = 1,2,\cdots$, N and P be probability measures on $\mathcal{A}$ such that $N_n \to N$ and P is absolutely continuous. Then $\|N_n * P - N * P\| \to 0$ as $n \to \infty$.

Proof. Let ν denote a left Haar measure on X. Let f be a version of $dP/d\mu$, and let g be a non negative continuous function with compact support. Put:

$$s(x) = \int f(y^{-1}x)N(dy); \qquad t(x) = \int g(y^{-1}x)N(dy)$$

$$s_n(x) = \int f(y^{-1}x)N_n(dy); \text{ and } \quad t_n(x) = \int g(y^{-1}x)N_n(dy.$$

Then:

$$s_n = \frac{dN_n * P}{d\nu}, \quad s = \frac{dN * P}{d\nu}, \quad \int t_n d\nu = \int t d\nu = \int g d\nu$$

Moreover $t_n(x) \to t(x)$ for each $x \in \chi$ so that by Scheffe's convergence theorem [11]

$$\int | t_n - t | d\nu \to 0 .$$

We get successively :

$$\begin{aligned}
\int |t - s| d\nu &= \int |\int g(y^{-1}x)N(dy) - \int f(y^{-1})N(dy)|\nu(dx) \\
&= \int |\int g(y^{-1}x) - f(y^{-1}x))N(dy)|\nu(dx) \\
&\leq \int \left[\int |g(y^{-1}x) - f(y^{-1}x)|N(dy)\right]\nu(dx) \\
&= \int \left[\int |g(y^{-1}x) - f(y^{-1}x)|\nu(dx)\right]N(dy) \\
&= \int \left[\int |g - f|d\nu\right]N(dy) = \int |f - g|d\nu
\end{aligned}$$

And similarly

$$\int |t_n - s_n|d\nu = \int |f - g|d\nu$$

Hence

$$\|N_n * P - N * P\| = \int |s_n - s|d\nu \leq \int |s_n - t_n|d\nu$$

$$+ \int |t_n - t|d\nu + \int |t - s|d\nu \leq 2\int |f - g|d\nu + \int |t_n - t|d\nu$$

so that

$$\lim_{n} \sup \| N \quad * P - N * P \| \leqq 2\int |f - g| d\nu \ .$$

The right hand side here may be made as small as we wish since the set of continuous functions with compact support is dense in $L_1(\nu)$.

Hence

$$\| N_n * P - N * P \| \to 0. \qquad \square$$

<u>Pooposition 11.</u>

Let X be a second countable locally compact abelian group with Borel class .

Let P_n; $n = 1,2,\cdots,M_n$; $n = 1,2,\cdots,N_n$; $n = 1,2,\cdots$ and $P << \mu$ be probability measures on $\mathcal{A}$ such that :

$$M_n * P_n \to P$$

$$\| N_n * P - P_n \| \to 0$$

Then $$\inf_{\Theta} \| P_{n,\Theta} - P \| \to 0$$

<u>Proof.</u>

1^{o}. Let us first prove the proposition under the additional assumptions that

$$P_n \to P$$

and that

$$N_n \to N$$

By proposition 10

$$\| N_n * P - N * P \| \to 0$$

so that

$$\|P_n - N * P\| \to 0$$

It follows that

$$N * P = P$$

Hence

$$\|P_n - P\| \to 0 .$$

2^0. Consider the general case. Put $\sigma_n = \inf_\theta \|P_{n,\theta} - P\|$ and let σ be a point of accumulation of σ_n; $n = 1,2,\cdots$ Let σ_{n_k}; $k = 1,2,\cdots$ be a sub sequence such that

$$\lim_{k\to\infty} \sigma_{n_k} = \sigma$$

By theorem 2.2 on page 59 in [10], M_n; $n = 1,2,\cdots$ and P_n;, $n = 1,2,\cdots$ are, respectively, right and left shift compact. It follows that we may - without loss of generality - assume that there is a sequence $a_1, a_2,\cdots$ so that $M_{n_k} * \delta_{a_k^{-1}} \to \tilde{M}$ and $\tilde{P}_k \underline{\underline{\text{definition}}} \delta_{a_k} * P_{n_k} \to \tilde{P}$. Hence $\tilde{M} * \tilde{P} = \lim(M_{n_k} * \delta_{a_k^{-1}} * \tilde{P}_k) = \lim M_{n_k} * P_{n_k} = P$. By assumption : $\|N_{n_k} * P - P_{n_k}\| \to 0$ so that $\|\delta_{a_k} * N_{n_k} * P - \tilde{P}_k\| \to 0$. We may - again without loss of generality - assume, since $\tilde{P}_k \to \tilde{P}$, that $\delta_{a_k} * N_{n_k} \to N$ where $N * P = \tilde{P}$.

This relation together with the relation $\tilde{M} * \tilde{P} = P$ implies $\Delta(P,\tilde{P}) = 0$ or equivalently that there is an a so that :

$\tilde{P} = \delta_a * P$. We have altogether shown that : $\tilde{M}_k * \tilde{P}_k \to \tilde{P}$, $\|\tilde{N}_k * \tilde{P} - \tilde{P}_k\| \to 0$, $\tilde{P}_k \to \tilde{P}$ and $\tilde{N}_k \to \tilde{N}$ where $\tilde{M}_k = \delta_a * M_{n_k} * \delta_{a_k}^{-1}$, $\tilde{N}_k = \delta_{a_k} * N_{n_k} * \delta_{a^{-1}}$ and $\tilde{N} = N * \delta_{a^{-1}}$. It follows from 1^o that $0 = \lim\|\tilde{P}_k - \tilde{P}\| = \lim\|\delta_{a_k} * P_{n_k} - \delta_a * P\| \geqq \lim \sigma_{n_k} = \sigma$ so that $\sigma = 0$. □

Example 12.

Let us - in two extreme cases - compare convergence as studied so far, with convergence for restrictions to finite sub parameter sets.

A sequence $\mathcal{E}_{P_1}, \mathcal{E}_{P_2}, \ldots$ converges for restrictions to finite sub parameter sets to $\mathcal{M}_i$ ($\mathcal{M}_a$) if and only if $\|P_{n\theta} - P_n\| \to 0$ ($\|P_{n\theta} - P_n\| \to 2$) for any θ ($\neq$ identity) in X. Ordinary convergence to $\mathcal{M}_i$ ($\mathcal{M}_a$), however, can never occur when X is not compact (when X is uncountable and the P_n's are absolutely continuous) since in this case $\delta(\mathcal{M}_i, \mathcal{E}_P) = 2$ for all P ($\delta(\mathcal{E}_{P_n}, \mathcal{M}_a) = 2$ for all n).

Corollary 13.

Suppose X is of the form $R^k Z^e$, that $P, P_1, P_2, \cdots$ are all symmetric, and that $P \ll \mu$. Assume $\inf_\theta \| P_{n,\theta} - P\| \to 0$ as $n \to \infty$. Then $\|P_n - P\| \to 0$ as $n \to \infty$.

Proof.

Let σ be a point of accumulation of $\sigma_n = \|P_n - P\|$; $n = 1,2,\cdots$ and let σ_{n_k}; $k = 1,2,\cdots$ be a sub sequence converging to σ. By assumption there is a sequence θ_k; $k = 1,2,\cdots$ in Θ so that

$$\|P_{n_k,\theta_k} - P\| = \|P_{n_k} - P_{-\theta_k}\| \to 0$$

and by symmetry

$$\|P_{n_k,-\theta_k} - P\| = \|P_{n_k} - P_{\theta_k}\| \to 0$$

Hence

$$\|P_{2\theta_k} - P\| \to 0$$

so that

$$\theta_k \to 0$$

Hence - since $P \ll \mu$ - $\sigma = \lim_k \sigma_{n_k} \leq \lim_k \sup \|P_{n_k} - P_{\theta_k}\| + \lim_k \sup \|P_{\theta_k} - P\| = 0$.

Remark.

That some conditions in addition to symmetry and absolute continuity are unavoidable may be seen from the following example. Put $X = \{0,1\}$ where $0+0 = 1+1 = 0$ and $0+1 = 1+0 = 1$. Put $P_1 = P_2 = \cdots , \delta$, $P = \delta_1$ and $\theta_1 = \theta_2 = \cdots = 1$. Then $\|P_{n,\theta_n} - P\| = 0$, $n = 1,2,\cdots$ but $\|P_n - P\| = 2$, $n = 1,2,\cdots$. Sequences $\{P_n\}$ and $\{\theta_n\}$ behaving like this may also easily be constructed on the circle group.

Let us next consider Cauchy sequences. It has been shown by LeCam [8] that - in general - Cauchy sequences of experiments for the pseudometric

$$\underline{\Delta}(\mathcal{E},\mathcal{F}) = \sup\{\Delta(\mathcal{E}_F, \mathcal{F}_F) \ ; \ F \text{ finite sub set of } \Theta\}$$

converges in that pseudometric. (If $\mathcal{E}$ and $\mathcal{F}$ are dominated experiments then $\underline{\Delta}(\mathcal{E},\mathcal{F}) = \Delta(\mathcal{E},\mathcal{F})$). It follows - for example - that to prove the completeness of $L_1(\mu)$ (considered as a set of translation experiments) it would suffice to show it is closed. This, however, seems to be easier said than done. We will not rely on this result here and instead demonstrate directly that Cauchy sequences (for Δ) of probability measures on $\mathcal{A}$ converges. This will then imply - as is easily seen - that $L_1(\mu)$ is closed and therefor complete.

Proposition 14.

Suppose $\sup_{m\geq n}\delta(P_m,P_n) \to 0$. Then there is a P so that $\delta(P_n,P) \to 0$.

Proof. We may - without loss of generality - assume that $\delta(P_{n+1},P_n) \leq 2^{-n}$; $n = 1,2,\cdots$. By assumption there is a sequence $M_1, M_2,\cdots$ of probability measures so that $\|M_{n+1} * P_{n+1}-P_n\| \leq 2^{-n}$. Hence: $\|M_1*\cdots* M_n * M_{n+1} * P_{n+1} - M_1*\cdots* M_n * P_n\| \leq 2^{-n}$. It follows that $M_1 *\cdots* M_n * P_n$ converges strongly. $\square$

Proposition 15.

Suppose $\sup_{m\geq n}\delta(P_m,P_n) \to 0$ and that $P_n \to P$. Then $\delta(P,P_n) \to 0$.

Proof. Put $\varepsilon_n = \sup_{m\geq n}\delta(P_m,P_n)$ and let $f \in C_0(X)^+$. By theorem 3 $P_n(f) \leqq \sup_x \int f(xy)P_m(dy) + \frac{1}{2}\varepsilon_n\|f\|$ when $m \geqq n$

$m \to \infty$ gives :

$$P_n(f) \leqq \sup_x \int f(xy)P(dy) + \frac{1}{2}\varepsilon_n\|f\| \ ; \quad f \in C_o(X)^+$$

so that

$$\delta(P,P_n) \leqq \varepsilon_n \ . \qquad \square$$

Theorem 16.

Suppose $\Delta(P_m,P_n) \to 0$ as $m,n \to \infty$ and that $P_n \to P$. Then $\Delta(P_n,P) \to 0$.

Proof. By the previous proposition $\delta(P,P_n) \to 0$. It remains to show that $\delta(P_n,P) \to 0$. Let $f \in C_o(X)^+$. Then

$$P_{n+m}(f) \leqq \sup_x \int f(xy)P_n(dy) + \frac{1}{2}\varepsilon_n\|f\| \qquad \text{where } \varepsilon_n \to 0.$$

$m \to \infty$ gives $\delta(P_n,P) \leqq \varepsilon_n$. $\square$

Theorem 17.

Suppose $\Delta(P_m P_n) \to 0$ as $m,n \to \infty$. Then there is a P so that $\Delta(P_n,P) \to 0$.

Remark : Clearly P is absolutely continuous if infinitely many of the P_n's are.

Proof of the theorem: By preposition 14 $\{P_n\}$ is left shift compact. The theorem now follows from theorem 16 . $\square$

Acknowledgement.

The author wishes to thank Professor Lucien LeCam for many helpful discussions.

6. References.

[1] Blackwell, D. (1951) Comparison of experiments, Proc. Second Berkeley symp. Math.Statist. Probab. Univ. of California Press, Berkeley and Los Angeles.

[2] Boll, C. (1955) Comparison of experiments in the infinite case. Ph.D. thesis, Stanford University.

[3] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann.Math.Statist. 23, 493-507.

[4] Day, M. (1961) Fixed point theorems for compact convex sets. Ill. J. Math. 5, pp.585-589.

[5] Greenleaf, F. (1969) Invariant means on topological groups Van Nostrand mathematical studies.

[6] Heyer, H. Personal communication.

[7] LeCam, L. (1964) Sufficiency and approximate sufficiency. Ann. Math. Statist. 35, 1419-1455.

[8] LeCam, L. Unpublished manuscript.

[9] Lehmann, E.L. (1959) Testing Statistical Hypotheses. Wiley, New York.

[10] Parthasarathy, K.R. (1967) Probability measures on metric spaces. Academic Press.

[11] Scheffé, H. (1947) A useful convergence theorem for probability distributions. Ann. Math. Statist. 18, 434-438.

[12] Strassen, V. (1965) The existence of probability measures with given marginals. Ann. Math. Statist. 36, 423.

[13] Torgersen, E.N. (1968) Comparison of experiments when the parameter space is finite. Ph.D.thesis, University of California, Berkeley. To be published in Z. Wahrscheinlichkeitstheorie und verw. Geb.

[14] Torgersen, E.N. Unpublished manuscripts. Abstracts. Ann.Math.Statist. (1969) 40, 727, 1155, 2219.